Speech diversity and speech interfaces - considering an inclusive future through stammering

Leigh Clark
Swansea University
l.m.h.clark@swansea.ac.uk

Benjamin R. Cowan
University College Dublin
benjamin.cowan@ucd.ie

Abi Roper
City, University of London
abi.roper.1@city.ac.uk

Stephen Lindsay
Swansea University
s.c.lindsay@swansea.ac.uk

Owen Sheers
Swansea University
o.g.sheers@swansea.ac.uk

ABSTRACT

The number of speech interfaces and services made available through them continue to grow. This has opened up interactions to people who rely on speech as a critical modality for interacting with systems. However, people with diverse speech patterns such as those who stammer are at risk of being negatively affected or excluded from speech interface interaction. In this paper, we consider what an inclusive speech interface future may look like for people who stammer. In doing so, we identify three key challenges: (1) developing effective speech recognition, (2) understanding the user experiences of people who stammer and (3) supporting speech interfaces designers through appropriate heuristics. We believe the interdisciplinary and cross-community strengths of venues like CUI are well positioned to address these challenges going forward.

CCS CONCEPTS

• **Human-centered computing** → **Accessibility technologies**; **Natural language interfaces**.

KEYWORDS

Stammer, stutter, inclusivity, accessibility, speech interface, speech diversity

ACM Reference Format:
Leigh Clark, Benjamin R. Cowan, Abi Roper, Stephen Lindsay, and Owen Sheers. 2020. Speech diversity and speech interfaces - considering an inclusive future through stammering. In *2nd Conference on Conversational User Interfaces (CUI '20), July 22–24, 2020, Bilbao, Spain.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3405755.3406139

1 A FUTURE - BUT FOR WHOM?

By 2023, intelligent personal assistants (IPAs) will be available on eight billion devices, including smart speakers, smartphones, wearables and smart televisions [26]. The increase in speech interfaces [5], and attempts to move towards more conversational interactions with users [6], point to speech becoming more common in the future. However, will such a speech driven future be open to all?

Currently, vital accessibility and inclusivity research on speech interface interaction has focused on older adult users [27], blind users [1] or those with limited hand dexterity [7]. For these users, speech is a critical and highly beneficial interaction modality, allowing them to interact with their devices where other modalities may prove more difficult. Yet there are some users groups that may be entirely excluded from the benefits derived from speech interface interaction. In particular, those with diverse speech patterns, who experience difficulties in fluent or typical speech production, may be unable to use current speech interfaces effectively. This paper focuses in particular on people who stammer and aims to highlight the need for significant work in developing speech technology experiences that do not exclude such users. We identify three key challenges in 1) providing effective speech recognition for people with diverse speech patterns; 2) understanding the key barriers and challenges faced by people who stammer when engaging with current speech interfaces and 3) supporting designers of speech interfaces with appropriate design heuristics.

2 STAMMERING & THE GROWTH OF SPEECH INTERFACES

Stammering[1] is a neurological condition characterised by disruptions to the "rhythmic flow of speech" [22]. These disruptions can include repetition, prolongation or hesitation of particular sounds or words. Estimates suggest a larger number of people stammer than previously thought - approximately 8% of children will stammer at some point and for up to to 3% of adults it will be a lifelong condition (up from 5% and 1% respectively) [28].

As speech interfaces become more mainstream, research is beginning to explore how they can be placed in a wide variety of contexts such as healthcare [14], automotive interfaces [13], education [12] and retail [21]. As a greater number of services become focused around speech, this may create a significant barrier for those with diverse speech patterns. This exclusion becomes particularly acute if any critical services are delivered through speech. While multimodal interfaces are sometimes available, users may still be faced with *hands-busy/eyes-busy* environments, additional user barriers, and the potential like of *choice* to interact with these systems.

[1] Also known as stuttering outside of the United Kingdom.

CUI '20, July 22–24, 2020, Bilbao, Spain

ACM ISBN 978-1-4503-7544-3/20/07…$15.00
https://doi.org/10.1145/3405755.3406139

3 CHALLENGES TO ADDRESS FOR PEOPLE WHO STAMMER

For people who stammer, using current speech interfaces may be challenging, excluding them from using speech as a modality. As speech interfaces becomes more commonplace this should be a concern for us in the conversational user interfaces (CUI) community. Current work on stammering in speech interaction is scant. We must therefore improve the volume of work to address this issue. We outline three key challenges and potential avenues for research on this topic to begin to address stammering user's experience.

3.1 Effective speech recognition

Automatic speech recognition (ASR) has improved drastically over the past decade [18], though there remain a plethora of speech signal variables that can negatively impact successful recognition [3]. Indeed, in accurately recognising dysarthic speech, recent research has shown there are still significant barriers for making ASR systems more inclusive, even with the move from generative models towards deep neural network (DNN) architectures [18].

Ongoing projects involving *Google AI* are focused on training speech recognition models on non-standard speech data. Project Euphonia[2] aims to collect samples on a wide variety of non-standard speech patterns, while Project Understood[3] focuses specifically on the speech patterns of people with Down's Syndrome.

While such projects are obviously welcome, there are still difficulties related to training data approaches. For people who stammer, it is common that features of stammering change over time [9] and across interactions [11]. This creates a challenge in gathering the volume of data required to create accurate models.

The volume of data potentially required to gather accurate ASR for users who stammer likely needs significant effort from large scale corporations who have the ability to gather such data. Yet it is important these models and data do not become proprietary, otherwise accessibility for these users will become monopolised. A solution is to adopt an open-source dataset approaches in the style of Mozilla Common Voice[4], LibriSpeech[5] and VoxForge[6], with the same objective on opening up ASR seen in closed sets.

Even with ASR training data available - how might an interaction look like in practice? We also have to consider ongoing technical challenges with ASR like *endpoint detection* - identifying when a speaker has finished speaking [15] - and how these processes may need to be altered with diverse speech patterns.

3.2 Understanding user experiences

Significant work in the CUI field has observed user's experiences with interfaces like IPAs, identifying issues such as the need to consider the potential gulf of expectation due to the humanness of such systems versus their actual functionality [8, 16, 19], the need to learn how to interact effectively [16] as well as how social and multiparty contexts impact the type of interactions we have with IPAs [8, 10, 24]. This work focuses almost entirely on users without significant accessibility requirements. Work on users with diverse speech is particularly scant - what key experiences do we need to understand in taking the first step towards an inclusive future?

We can learn from workshops on making speech interfaces accessible [4] and ensure people with diverse speech patterns are also included in the future of interaction research. Additionally, we can adopt methods like *participatory design* and *co-design* in including people with diverse speech patterns in the research process. This can help address any gulf in experience between designers and the users they are designing for. Consequently, we may find design decisions that transcend language in positively impacting user experiences - something that would go beyond language-limited speech data. We can also engage with charities[7] and institutions to build inclusive networks to support this research.

We can also consider the potential benefits speech interfaces may offer people with diverse speech patterns. IPAs and other ASR systems have been touted as methods for creating forms of speech therapy (e.g. [23, 25]). Speech interfaces may be able to follow existing research on supporting people who stammer (e.g. [17]).

3.3 Supporting designers through heuristics

There is interest in developing heuristics for speech interface designers [20], though how to design for people with stammers remains unclear. We may need to develop heuristics that can be implemented in systems that talk to people who stammer. While we have advice for talking to people who stammer (e.g. [29]), this may not be feasible or transferable to human-computer interaction (HCI).

Showing users that an interface is listening is often supported multimodally in speech interface interaction, through audible notifications or visual indicators (e.g. Amazon Echo's ring [2]). Providing time for people on phone calls may be applicable for interactive voice response (IVR) systems, while considerations of eye contact would be reserved for embodied areas of HCI and robotics.

It is difficult to envisage speech interface designers implementing requests that users slow down their speech or relax, at least for general use systems like IPAs. Interruptions and attempts at guessing or finishing the words of people who stammer may not be appropriate in speech interfaces. Conversely, the fundamental differences of speaking with machines and speaking with people (e.g. [6]) may mean these do not carry the same social weight. This ambiguity again requires an understanding of user experiences in order to develop appropriate design heuristics and understand how they may be altered depending on the context of interaction.

4 CONCLUSION

It is an exciting time for speech interfaces and the expansion of services and interactions available through them. However, we must consider the significant number of people with diverse speech patterns such as stammering. In considering stammering, we have outlined three crucial challenges in developing effective ASR, understanding user experiences of people who stammer and creating appropriate heuristics to support speech interface designers. We must consider what an inclusive future looks like for people with diverse ranges of speech patterns such as stammering and utilise the cross-community strengths in venues like CUI to do so.

[2] https://sites.google.com/view/project-euphonia
[3] https://projectunderstood.ca
[4] https://voice.mozilla.org
[5] https://www.openslr.org/12
[6] http://www.voxforge.org
[7] https://stamma.org

REFERENCES

[1] Ali Abdolrahmani, Ravi Kuber, and Stacy M Branham. 2018. " Siri Talks at You" An Empirical Investigation of Voice-Activated Personal Assistant (VAPA) Usage by Individuals Who Are Blind. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*. 249–258.

[2] Amazon. 2020. What Do the Lights on Your Echo Device Mean? https://www.amazon.com/gp/help/customer/display.html?nodeId=GKLDRFT7FP4FZE56 Accessed 23rd Feb 2020.

[3] Mohamed Benzeghiba, Renato De Mori, Olivier Deroo, Stephane Dupont, Teodora Erbes, Denis Jouvet, Luciano Fissore, Pietro Laface, Alfred Mertins, Christophe Ris, et al. 2007. Automatic speech recognition and speech variability: A review. *Speech communication* 49, 10-11 (2007), 763–786.

[4] Robin N. Brewer, Leah Findlater, Joseph "Jofish" Kaye, Walter Lasecki, Cosmin Munteanu, and Astrid Weber. 2018. Accessible Voice Interfaces. In *Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '18)*. Association for Computing Machinery, New York, NY, USA, 441–446. https://doi.org/10.1145/3272973.3273006

[5] Leigh Clark, Philip Doyle, Diego Garaialde, Emer Gilmartin, Stephan Schlögl, Jens Edlund, Matthew Aylett, João Cabral, Cosmin Munteanu, Justin Edwards, et al. 2019. The State of Speech in HCI: Trends, Themes and Challenges. *Interacting with Computers* 31, 4 (2019), 349–371.

[6] Leigh Clark, Nadia Pantidi, Orla Cooney, Philip Doyle, Diego Garaialde, Justin Edwards, Brendan Spillane, Emer Gilmartin, Christine Murad, Cosmin Munteanu, et al. 2019. What makes a good conversation? challenges in designing truly conversational agents. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.

[7] Eric Corbett and Astrid Weber. 2016. What can I say? addressing user experience challenges of a mobile voice user interface for accessibility. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services*. 72–82.

[8] Benjamin R Cowan, Nadia Pantidi, David Coyle, Kellie Morrissey, Peter Clarke, Sara Al-Shehri, David Earley, and Natasha Bandeira. 2017. "What Can I Help You With?": Infrequent Users' Experiences of Intelligent Personal Assistants. *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services - MobileHCI '17* (2017), 1–12. https://doi.org/10.1145/3098279.3098539

[9] Isobel Crichton-Smith. 2002. Communicating in the real world: Accounts from people who stammer. *Journal of fluency disorders* 27, 4 (2002), 333–352.

[10] Joel E Fischer, Stuart Reeves, Martin Porcheron, and Rein Ove Sikveland. 2019. Progressivity for voice interface design. In *Proceedings of the 1st International Conference on Conversational User Interfaces*. 1–8.

[11] Rosemarie Hayhow, Anne Marie Cray, and Pam Enderby. 2002. Stammering and therapy views of people who stammer. *Journal of fluency disorders* 27, 1 (2002), 1–17.

[12] Hyunhoon Jung, Hee Jae Kim, Seongeun So, Jinjoong Kim, and Changhoon Oh. 2019. TurtleTalk: an educational programming game for children with voice user interface. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–6.

[13] Andrew L Kun et al. 2018. Human-machine interaction for vehicles: Review and outlook. *Foundations and Trends® in Human–Computer Interaction* 11, 4 (2018), 201–293.

[14] Liliana Laranjo, Adam G Dunn, Huong Ly Tong, Ahmet Baki Kocaballi, Jessica Chen, Rabia Bashir, Didi Surian, Blanca Gallego, Farah Magrabi, Annie YS Lau, et al. 2018. Conversational agents in healthcare: a systematic review. *Journal of the American Medical Informatics Association* 25, 9 (2018), 1248–1258.

[15] Qi Li, Jinsong Zheng, Augustine Tsai, and Qiru Zhou. 2002. Robust endpoint detection and energy normalization for real-time speech and speaker recognition. *IEEE Transactions on Speech and Audio Processing* 10, 3 (2002), 146–157.

[16] Ewa Luger and Abigail Sellen. 2016. "Like Having a Really Bad PA": The Gulf between User Expectation and Experience of Conversational Agents. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16* (2016), 5286–5297. https://doi.org/10.1145/2858036.2858288

[17] Roisin McNaney, Christopher Bull, Lynne Mackie, Floriane Dahman, Helen Stringer, Dan Richardson, and Daniel Welsh. 2018. StammerApp: Designing a Mobile Application to Support Self-Reflection and Goal Setting for People Who Stammer. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–12.

[18] Meredith Moore, Hemanth Venkateswara, and Sethuraman Panchanathan. 2018. Whistle-blowing ASRs: Evaluating the need for more inclusive automatic speech recognition systems. In *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, Vol. 2018. 466–470.

[19] Roger K Moore. 2017. Is spoken language all-or-nothing? Implications for future speech-based human-machine interaction. In *Dialogues with Social Robots*. Springer, 281–291.

[20] Christine Murad, Cosmin Munteanu, Benjamin R Cowan, and Leigh Clark. 2019. Revolution or Evolution? Speech Interaction and HCI Design Guidelines. *IEEE Pervasive Computing* 18, 2 (2019), 33–45.

[21] Christi Olson and Kelli Kemery. 2019. Voice report: From answers to action: customer adoption of voice technology and digital assistants. *Microsoft Search and Market Intelligence, Tech. Rep* (2019).

[22] World Health Organization. 2010. ICD-10 Version:2010. http://apps.who.int/classifications/icd10/browse/2010/en#/F98.5 Accessed 23rd Feb 2020.

[23] Rebecca Palmer, Pam Enderby, and Mark Hawley. 2007. Addressing the needs of speakers with longstanding dysarthria: computerized and traditional therapy compared. *International journal of language & communication disorders* 42, S1 (2007), 61–79.

[24] Martin Porcheron, Joel E Fischer, Stuart Reeves, and Sarah Sharples. 2018. Voice interfaces in everyday life. In *proceedings of the 2018 CHI conference on human factors in computing systems*. 1–12.

[25] Alisha Pradhan, Kanika Mehta, and Leah Findlater. 2018. " Accessibility Came by Accident" Use of Voice-Controlled Intelligent Personal Assistants by People with Disabilities. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–13.

[26] Juniper Research. 2018. Digital Voice Assistants in Use to Triple to 8 Billion by 2023, Driven by Smart Home Devices. shorturl.at/dgoGL. Accessed 22nd Feb 2020.

[27] Sergio Sayago, Barbara Barbosa Neves, and Benjamin R Cowan. 2019. Voice assistants and older people: some open issues. In *Proceedings of the 1st International Conference on Conversational User Interfaces*. 1–3.

[28] Stamma. 2020. Stammer in the Population. https://stamma.org/news-features/stammering-population Accessed 23rd Feb 2020.

[29] Stamma. 2020. Talking With Someone Who Stammers. https://stamma.org/about-stammering/talking-someone-who-stammers Accessed 23rd Feb 2020.